

*Abstract of the Disclosure*

An arrangement for inserting advertisement media into a streaming multimedia file destined for an end-user uses a caching inserter including a first cache for storing a loaded advertisement file and a second cache for holding the streaming multimedia file  
5 that needs to be buffered as the advertisement file is sent to the end-user. The advertisement file may be inserted at the beginning of the multimedia file, the end, or at any point during the stream. Once the advertisement begins to be sent to the end-user, the incoming multimedia stream is cached and then is played out of the cache once the advertisement has ended. As a result of the time delay associated with the advertisement,  
10 the remainder of the multimedia file must be cached then played out until the stream is complete.